Transcrição

Começaremos vendo uma maneira de **salvarmos** o notebook que acabamos de criar para acompanharmos as aulas.

Para isso, bastará acessarmos "File > Download .ipynb" para fazermos o download do arquivo, e em seguida o salvare

Logo, recomendamos salvar o projeto ao final de cada aula, para então podermos fazer seu upload e o de `dados.csv` ,

Já temos o dataset e DataFrame em nosso notebook. Com o método `head()` do Pandas, visualizaremos somente os ci

```
dados.head()
```

Os **tipos de dados** devem ser bem entendidos, pois cada um deles possui um tipo de estatística e tratamento da inform

Classificaremos os dados entre basicamente dois tipos: **qualitativos** e **quantitativos**. O primeiro expressa uma **qualid**

Entenderemos a partir do nosso DataFrame; `UF` é bastante simples de classificarmos como qualitativo, pois não é uma

Já `Anos de Estudo` é uma variável de dados qualitativos também, afinal cada código numérico representa uma classe exemplo.

Cada classificação pode ser dividida em duas categorias. No caso dos qualitativos, teremos os **ordinais** e **nominais**.

Já os quantitativos se dividem em **discretos** e **contínuos**. Veremos exemplos de cada um mais adiante com o uso do Py

Começando pelas variáveis qualitativas ordinais, não identificaremos as `Sexo` , `Cor` e `UF` como tais, pois não poden

Já no caso de `Anos de Estudo` , conseguiremos fazer uma **ordenação**, então poderemos classificá-la desta forma. Cas simples na parte de "1.2 Variáveis qualitativas ordinais" no notebook.

```
dados['Anos de Estudo']
```

Para vermos apenas os valores únicos, aplicaremos `.unique()` ao comando da célula.

```
dados['Anos de Estudo'].unique()
```

O resultado será um `array()` fora de ordenação. Para ordenarmos os dados, adicionaremos a *building function* do Py

```
sorted(dados['Anos de Estudo'].unique())
```

Com isso, visualizaremos a lista de valores do dicionário de `Anos de Estudo` . Logo, comprovaremos que se trata de u

Abordando as qualitativas nominais, teremos `Sexo`, `Cor` e `UF`. As imprimiremos na célula desta parte do notebook. corretas.

```
sorted(dados['Sexo'].unique())
```

```
sorted(dados['Cor'].unique())
```

```
sorted(dados['UF'].unique())
```

Desta forma, veremos a numeração do código atribuída pelos profissionais do IBGE.

Estes valores não podem ser hierarquizados para serem ordenados, o qua as classificam como qualitativas nominais.

Mais adiante, abordaremos as variáveis quantitativas discretas. Poderemos classificar a `Idade` de diversas formas, de **inteiros**.

Quando a pesquisa é feita pelo IBGE, os entrevistadores perguntam quantos anos completos a pessoa tem, sem contar

Na parte homônima do nosso notebook, escreveremos `dados` com `Idade` que pode ser aplicada em diversas operaçõ

```
dados.Idade.min()
```

Com isso, veremos a idade mínima presente em nosso dataset. Também poderemos exibir a idade máxima com `max()`

Faremos uma impressão com `print()` recebendo `'De %s até %s anos'`, seguido de `% ()` contendo os dois últimc

```
print('De %s até %s anos' % (dados.Idade.min(), dados.Idade.max()))
```

Rodando este código na célula desta parte do notebook, veremos a impressão de `De 13 até 99 anos`. Portanto, esta

Também poderia ser quantitativa contínua, pois é possível representarmos idades exatas que computem os meses e dia

A variável `Idade` seria qualitativa ordinal se fosse inserida em intervalos, como em casos onde faixas etárias são rele

Passando para a abordagem das variáveis quantitativas contínuas, já citamos que a `Idade` poderia ser deste tipo em al

Como a `Renda` veio da fonte do IBGE sem a contagem de centavos, criamos a variável `Altura` justamente para vern

Logo, esta última é classificada como quantitativa contínua. Aplicaremos o mesmo comando anterior na célula desta p

```
print('De %s até %s metros' % (dados.Altura.min(), dados.Altura.max()))
```

Com isso, veremos o intervalo entre o valor mínimo e o máximo da variável `Altura` medido em metros.

Se escrevermos `['Altura']` ao invés de somente `Altura` para pegarmos seus `dados`, o resultado será igual.

```
print('De %s até %s metros' % (dados['Altura'].min(), dados['Altura'].max()))
```

Esta outra maneira de escrever é útil para os casos em que uma variável possui mais de uma palavra separada por espa
`Estudo']` na operação como já fizemos anteriormente.

Mais adiante em nosso notebook, encontraremos um esquema gráfico que traduz de forma simples em um diagrama to

Em resumo, teremos a variável qualitativa ordinal `Anos de Estudo` e as qualitativas nominais `Sexo`, `Cor` e `UF`, be

Com isso, avançaremos para o estudo da **Distribuição de Frequências**; a seguir, usando a biblioteca Pandas, aprender